Pre-LUCA Protein Evolution Reconstructed from an Imprint Conserved within the Genetic Code

Brian K. Davis :: Research Foundation of Southern California
8861 Villa La Jolla Dr. #13595, La Jolla, CA 92037, USA **::** davis@resfdnsca.org

Introduction: Recent insights into the formation of the genetic code[1], around 4×10^9 years ago, have made it possible to reconstruct protein evolution prior to the Last Universal Common Ancestor (LUCA). The time-order of pre-LUCA protein formation being inferred from the 'code age' of the ancestral residue profile.

Procedure: Antecedents of the H^+-ATPase proteolipid helix-1 subunit (PL-h1) and ferredoxin (Fd) were identified by a neighbor-joining analysis of sequences from a database search[2]; the latter protein was then reconstructed.[3] From the distribution of residue synthesis path-lengths, the 'code age' of each pre-LUCA protein was determined. In addition, residue hydropathy free energy (ΔG_T), from aqueous/non-aqueous phase (dielectric constant, 2) partitioning, charge, site mutability, and function were appraised.

Results:

☐ GENETIC CODE FORMED IN THREE STAGES

Successive recruitment of the codon 5'-, mid-, and 3'-base produced three distinct stages of code formation. Initially, both diacid amino acids (Asp[1], Glu[1]), and their amides (Asn[2], Gln[2]), together with a termination signal (Ter), acquired 5'-base specific codons within the NAN column of triplets. They form a hydrophilic cluster of short path (1 or 2 step) amino acids, produced by oxaloacetate, or oxoglutarate, reacting with one, or two, ammonium ions.

Columnwise expansion, through mid-base recruitment, added ten amino acids, formed on 2- to 8-step paths. All are hydrophobic, discounting Pro[4], and those in the NUN set form a highly hydrophobic cluster. As this set has the longest mean synthesis steps, nearly 7 steps, a hydrophobic attractor evidently drove encoding of mid-stage amino acids.

3'-Base recruitment placed six basic, aromatic, and hetero-cyclic amino acids (9-14 steps) in split-boxes with homologs.

NH₄⁺ Fixers		Expansion						Overprint
5↓ mid→		C		G		U		3↓
	ΔG_T		ΔG_T		ΔG_T		ΔG_T	
G	Asp¹ 9.2 / Glu¹ 8.2	Ala²	-1.6	Gly⁵	-1.0	Val⁵	-2.6	Y / R
A	Asn² 4.8 / Lys¹⁰ 8.8	Thr⁶	-1.2	Ser⁴ -0.6 / Arg⁹ 12.3		Ile⁷ -3.1 / Met⁷ -3.4		Y / A / G
C	His¹³ 3.0 / Gln² 4.1	Pro⁴	0.2	Arg⁹	12.3	Leu⁸	-2.8	Y / R
U	Tyr¹¹ 0.7 / Ter	Ser⁴	-0.6	Cys⁵ -2.0 / Ter / Trp¹⁴ -1.9		Phe¹¹ -3.7 / Leu⁸ -2.8		Y / A / G
	Hydrophilic cluster	**Transitional boxes**				**Hydrophobic cluster**		

☐ $\Delta G_T > 0$ kcal/mol., hydrophilic   Superscript, synthesis path-length
☐ $0 > \Delta G_T > -2.0$ kcal/mol., hydrophobic
☐ $-2.0 \geq \Delta G_T$ kcal/mol., strongly hydrophobic

☐ CODE EXPANSION TOWARD A HYDROPHOBIC ATTRACTOR

Occurrence of a linear relation between amino acid synthesis path-length and residue transfer free energy, during expansion from the NH_4^+ Fixers Code, establishes hydrophobicity as the principal determinant of amino acid addition to the code. A dynamic interpretation of the code clusters results. The hydrophilic cluster with both diacid amino acids and their amides, formed in 1 to 2 steps, together with chain termination triplets, in the NAN column, formed the first code. Mid-base codon assignments to increasingly hydrophobic amino acids followed, leading to the cluster of highly hydrophobic amino acids, synthesized on 5 to 8 step paths, at weak-bonding NUN triplets.

$\Delta G_T = 7.0 - 1.5 L$
$r = 0.84$

☐ PROTEINS AS MARKERS IN PRE-LUCA EVOLUTION OF THE CELL

All 11 residues in pre-LUCA PL-h1, a trans-'membrane bilayer' subunit of H^+-ATPase, were hydrophobic: I⁷G⁵A²G⁵L⁸A²V⁵G⁵L⁸A²A².[2] Four highly hydrophobic residues, formed on 5- to 8-step paths, place the origin of both PL-h1 and that of a fuctional cell membrane at the final stage of code expansion. Consistent with this, markedly different kinds of amino acids, having a large basic, or cyclic, side-chain, were subsequently added to the code, through recruitment of the codon 3'-base, in its final stage.

With residue path-lengths up to 5-steps, pre-LUCA Fd antecedent, pro-Fd[5], has a mid-expansion stage 'code age'. Thus, it evidently predates PL-h1 and the cell membrane. Attachment of a negatively charged N-terminal 7-residue segment, with 2 Asp[1] and 3 Glu[1], to the 16 residue [4Fe-4S] cofactor-binding segment, moreover, furnishes evidence that pro-Fd[5] anchored its cofactor, by charge attraction, to a cationic mineral surface, in a pre-cell metabolic system.[2,3]

Inclusion of 14 hydrophobic residues - with 9 highly hydrophobic - in the 16 residue cofactor-binding segment, furthermore, would prolong pro-Fd-[5] attachment (Fajan-Paneth effect) to a cationic mineral surface, supporting a pre-cell metabolism.
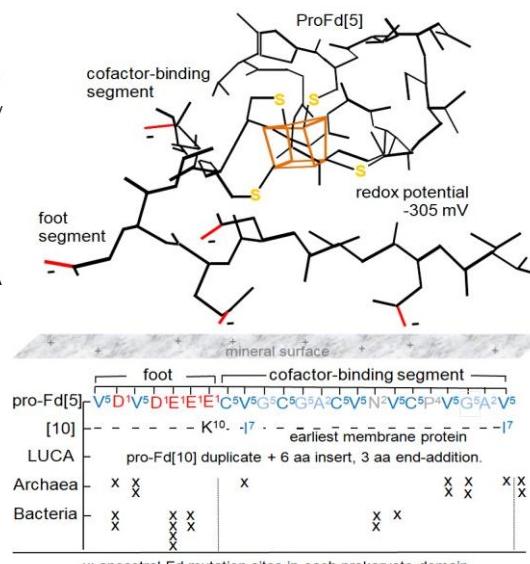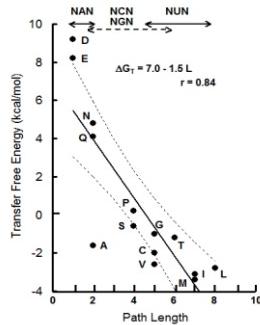
Ancestral Archaea and Bacteria Fd both differ from LUCA Fd by mutations at 11 sites, placing them equi-distant from their common ancestor. Whereas the Archaean Fd has a single mutation at the site of a charged residue within its 7-residue foot, Bacteria Fd has 8 mutations at charged sites within its foot:
$\chi^2_1 = 6.77$; $p = 9.28 \times 10^{-3}$ (corrected for continuity). A significantly slower rate of Archaean transition from the surface-binding form of Fd implies longer retention of the transitional surface/cell intermediate state.

Residue path-distances of the transmembrane PL-h1 protein placed cell membrane formation at the end of codon mid-base recruitment, during evolution of the universal genetic code. Thus, the transition from a mineral surface system to membrane-bound cells significantly preceded LUCA.

ProFd[5]
cofactor-binding segment
foot segment
redox potential −305 mV
mineral surface

	foot	cofactor-binding segment
pro-Fd[5]	V⁵D¹V⁵D¹E¹E¹E¹	C⁵V⁵G⁵C⁵G⁵A²C⁵V⁵N²V⁵C⁵P⁴V⁵G⁵A²V⁵
[10]	– – – K¹⁰ – I⁷	– – – – – – – – I⁷
LUCA		pro-Fd[10] duplicate + 6 aa insert, 3 aa end-addition.

earliest membrane protein

Archaea: x x / x x / x x x / x|x x
Bacteria: x x / x x x x / x x

x: ancestral Fd mutation sites in each prokaryote domain.

Conclusion: Pre-LUCA Fd and H^+-ATPase PL-h1 residue sequence identfied cofactor-anchoring to mineral surface and membrane formation as hydrophobic attractors in codon mid-base recruitment during expansion stage of genetic code formation.

References: 1. Davis BK (2020) Genetic code evolution reconstructed with aligned metrics. *Biorxiv doi:10.1101/2020.07.29.22778*
2. Davis BK (2002) Molecular evolution before the origin of species. *Prog. Biophys. Mol. Biol. 79*, 77-133.
3. Nørgaard H (2009) Characterization of ancient ferredoxins & Chlamydia Trachomatis ribonucleotide reductase. Ph.D. Thesis, Technical University of Denmark, Lyngby. https://orbit.dtu.dk/files/5050569/Hanne%20Nørgaard.pdf.
Nørgaard H ,Helt SS, Ooi BL, Hagen WR, Christensen HEM 2009. *J Biol Inorg Chem* 14 Supp.1, 212, P581.